

Санкт-Петербургский государственный университет
Кафедра компьютерного моделирования и многопроцессорных
систем

Чернышов Дмитрий Олегович

Выпускная квалификационная работа бакалавра

**Оценка оптимальной стоимости платных
услуг на сайте Avito**

Направление 010400

Прикладная математика, фундаментальная информатика
и основы программирования

Научный руководитель,
доктор техн. наук,
профессор
Буре В.М.

Санкт-Петербург

2016

Содержание

Введение	3
Обзор литературы	5
Глава 1. Сайт Avito	7
1.1. История сайта	7
1.2. Статистика сайта	8
1.3. Сервисы и услуги	9
Глава 2. Задача установления стоимости платных услуг	13
2.1. Постановка задачи	13
2.2. Методы кластеризации	15
2.3. Эмпирическая оценка стоимости платных услуг	19
2.4. Основные результаты и обсуждение	20
Глава 3. Математическая модель репутации пользователей	27
3.1. Постановка задачи	27
3.2. Метод оценки репутации	28
3.3. Основные результаты и обсуждение	29
Заключение	34
Список литературы	35
Приложение	38

Введение

Данная работа посвящена решению двух задач. Первая из них – определения цен на платные услуги сайта Avito. Эта задача актуальна на сегодняшний день по причине того, что в данный момент на ресурсе существуют неточности в значениях стоимости услуг. Представленный метод решения позволяет пересчитать эти цены с учетом благосостояния каждого региона и дополнительных характеристик каждой категории объявлений. Метод состоит из двух этапов. Первый этап – анализ данных о благосостоянии населения регионов Российской Федерации и проведение процессов кластеризации. Второй этап – проведение опроса пользователей и построение эмпирической оценки функции спроса на вышеуказанные услуги и определении оптимальной цены на них. Статья о разработке данного метода была успешно представлена на международной научной конференции аспирантов и студентов «Процессы управления и устойчивость» Control Processes and Stability (CPS'16).

Вторая задача – построение модели репутации пользователей. На данный момент сайт Avito является свободной площадкой для размещения объявлений и не участвует в процессе сделки между пользователями. Это отличает сайт от популярных торговых площадок как, например, eBay. В 2015 году компания

Avito приобрела сервис доставки CheckOut, который агрегирует в себе различные способы доставок. Указанный сервис позволит контролировать процесс сделки и, следовательно, получать отзывы о продавцах. В связи с этим, появится необходимость во введении рейтинга продавцов, который будет показывать надежность продавца. В предложенной модели минимизировано влияние искусственного завышения или занижения репутации мошенниками или конкурентами.

Обзор литературы

В первой части работы, для решения первой задачи, использовался метод кластеризации данных. Такой подход к решению выбран в статье [1]. Данные для кластеризации брались из базы федеральной статистики ЕМИСС. Для всех процессов использовались данные о средней заработной плате [2] и численности населения [3]. В зависимости от категории объявлений, в качестве дополнительных входных параметров были использованы данные о количестве поставленных на учет автомобилей [4], средней стоимости квадратного метра жилья [5], количестве зарегистрированных организаций [6] и объеме предоставленных платных услуг населению [7]. Существующие методы кластеризации описаны в книге [8]. Процессы кластеризации проводились в среде MATLAB с помощью функции *k-means*, которая реализована в пакете Statistics Toolbox [10]. После проведения процессов кластеризации производилось построение эмпирической оценки функции спроса, принцип которого описан в книге [9]. Во второй части работы, для решения второй задачи использовались статьи, в которых представлена информация о доверии [11] и выборе продавца [12] в области электронной коммерции. Построение модели репутации изложено в статье [13], в которой описываются нюансы и принци-

пы работы системы рейтинга основных торговых площадок. На основе данной статьи предложена новая, адаптированная для сайта Avito, математическая модель.

Глава 1. Сайт Avito

Avito – один из крупнейших сайтов объявлений в России. На данный момент на активно более 30 миллионов объявлений и миллионы людей ежедневно пользуются этим сервисом. Мобильные технологии позволяют обращаться к этой торговой площадке из любого места, а недавно запущенные сервисы привлекают все больше рекламодателей.

1.1. История сайта

Сайт был создан 15 октября 2007 года. Специализировался на товарах повседневного спроса. Помимо публикации объявлений, существовала возможность размещения аукционов, но в 2009 от этой функции было решено отказаться, вместо этого все внимание руководства компании сфокусировалось на работе с объявлениями. В том же году на сайте появился сервис «Магазины», позволяющий размещать на сайте витрины для товаров.

- В 2011 году запустились мобильная версия сайта и мобильное приложение.
- В 2012 году появились категории «Авто» и «Недвижимость».

- В 2013 году произошло слияние Avito и ресурсов OLX.ru и Slando.ru. Запустилась новая категория «Услуги». В этом же году состоялся запуск проекта Domofond.ru.
- В 2015 году состоялся запуск сервиса контекстной рекламы «Avito Контекст».

1.2. Статистика сайта

- В 2010 году количество уникальных посетителей достигло отметки 10 млн в месяц.
- В 2012 году количество уникальных посетителей достигло отметки 10 млн в неделю.
- В 2014 году компания заняла 10-е место в рейтинге крупнейших интернет-компаний в России по версии Forbes.
- В 2015 году ежемесячная посещаемость ресурса со стационарных устройств составила 25 млн человек, согласно статистике TNS Web Index report. Также в этом году ежемесячная аудитория находилась на уровне 35 млн уникальных посетителей, из которых 10,1 млн человек использовали мобильные устройства, согласно отчету Vostok New Ventures.

- Avito входит в ТОП-5 крупнейших сайтов России по версии Liveinternet, а также является наиболее посещаемым российским сайтом в категории «Товары и услуги».

1.3. Сервисы и услуги

В данный момент на сайте Avito существуют 10 категорий объявлений: «Авто», «Недвижимость», «Работа», «Услуги», «Личные вещи», «Для дома и дачи», «Бытовая электроника», «Хобби и отдых», «Животные», «Для бизнеса». На сайте предлагаются как новые, так и товары, бывшие в употреблении. Сайт имеет несколько сервисов для бизнеса:

- Магазины – сервис, позволяющий открыть на сайте собственную страницу с магазином. Возможность использования как компанией, так и частными предпринимателями.
- Промо – сервис, позволяющий создать рекламу, направленную на конкретную аудиторию сайта.
- Контекст – сервис, привлекающий пользователей с Avito на страницы интернет-магазинов с помощью контекстной рекламы.
- CheckOut – логистический сервис, агрегирующий услуги различных служб доставки.

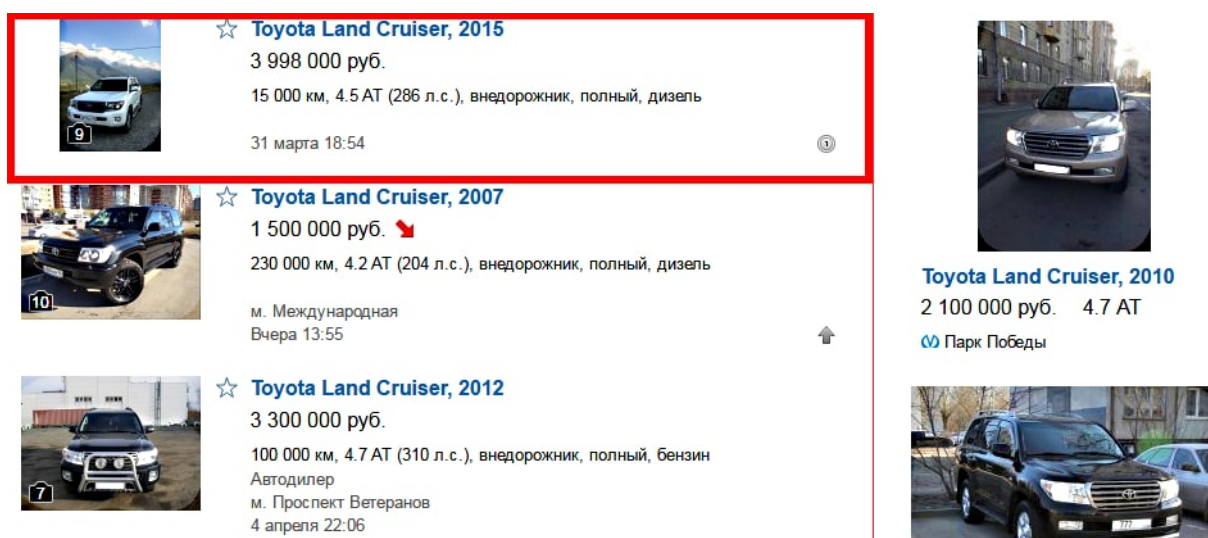
- ActiAgent – инструмент для работы агентов и брокеров.

Позволяет управлять своей базой, размещать и отслеживать объявления на площадках, вести клиентов.

Помимо представленных сервисов для бизнеса, пользователям предоставляются дополнительные платные услуги. Все они направлены на привлечение внимания потенциальных покупателей теми или иными способами.

Премиум-размещение

Премиум-объявления отображаются сверху каждой страницы списка результатов поиска и обозначаются соответствующим значком (см. рис. 1). Премиум-объявления упорядочены по дате или цене и соответствуют поисковому запросу, региону и выбранной категории.



The screenshot displays a list of car advertisements. The first three listings are enclosed in a red rectangular border, indicating they are premium placements. Each listing includes a car image, a star icon, the model name and year, price, specifications, and location. The first listing is for a 2015 Toyota Land Cruiser, the second for a 2007 model, and the third for a 2012 model. To the right, two more listings are visible: a 2010 Toyota Land Cruiser and another 2010 model.






Image	Model	Price	Specs	Location
	Toyota Land Cruiser, 2015	3 998 000 руб.	15 000 км, 4.5 AT (286 л.с.), внедорожник, полный, дизель	31 марта 18:54
	Toyota Land Cruiser, 2007	1 500 000 руб.	230 000 км, 4.2 AT (204 л.с.), внедорожник, полный, дизель	м. Международная Вчера 13:55
	Toyota Land Cruiser, 2012	3 300 000 руб.	100 000 км, 4.7 AT (310 л.с.), внедорожник, полный, бензин	Автодилер м. Проспект Ветеранов 4 апреля 22:06
	Toyota Land Cruiser, 2010	2 100 000 руб.	4.7 AT	Парк Победы
				

Рис. 1: Премиум-объявление

VIP-объявление

VIP-объявления располагаются на странице результатов поиска в специальном блоке и обозначаются соответствующим значком (см. рис. 2). В нем могут располагаться не более 3 VIP-объявлений, поэтому они показываются в случайном порядке. Все VIP-объявления соответствуют поисковому запросу пользователя, региону и категории.

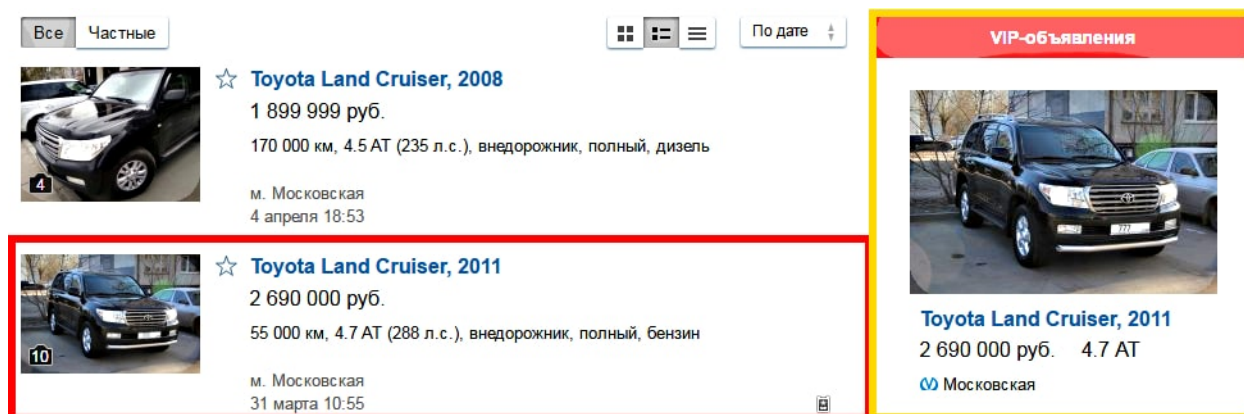




Рис. 2: VIP-объявление

Поднятие объявления в поиске


Поднятие объявления в поиске – услуга, с помощью которой объявление дважды поднимается на верхнюю позицию в списке результатов поиска. Также обозначается соответствующим значком (см. рис. 3). Второе поднятие происходит через 24 часа после первого.




☆ **Toyota Land Cruiser, 2007**
1 500 000 руб. 🚩
230 000 км, 4.2 АТ (204 л.с.), внедорожник, полный, дизель
м. Международная
Вчера 13:55




☆ **Toyota Land Cruiser, 2012**
3 300 000 руб.
100 000 км, 4.7 АТ (310 л.с.), внедорожник, полный, бензин
Автодилер
м. Проспект Ветеранов
4 апреля 22:06



☆ **Toyota Land Cruiser, 2011**
2 590 000 руб.
40 000 км, 4.5 АТ (235 л.с.), внедорожник, полный, дизель
4 апреля 19:07



Toyota Land Cruiser, 2010
2 100 000 руб. 4.7 АТ
Парк Победы




Toyota Land Cruiser, 2011
2 690 000 руб. 4.7 АТ
Московская


Рис. 3: Поднятие объявления в поиске

Выделение объявления


Выделенные объявления помечаются специальной иконкой и показываются на золотом фоне в общем списке на странице результатов поиска (см. рис. 4). В качестве бонуса после применения услуги объявление однократно поднимается в поиске.



☆ **Toyota Hilux, 2011**
1 250 000 руб.
95 000 км, 3.0 АТ (170 л.с.), пикап, полный, дизель
Автодилер
м. Ладжская
3 апреля 16:13



☆ **Toyota Hilux, 2012**
1 370 000 руб.
110 000 км, 3.0 АТ (170 л.с.), пикап, полный, дизель
Автодилер
м. Лесная
2 апреля 20:26



Volkswagen Multivan, 2013
2 199 000 руб. 2.0 АМТ
Проспект Ветеранов

Рис. 4: Выделение объявления

Глава 2. Задача установления стоимости платных услуг

2.1. Постановка задачи

Существование платных услуг порождает задачу установления цен на эти услуги. Функционал сайта позволяет публиковать объявления в любом регионе Российской Федерации. В каждой области различный уровень жизни и благосостояния населения. Кроме того, на сайте публикуются как товары, отдающиеся даром, так и товары, стоимостью несколько сотен миллионов рублей. Это должно приниматься во внимание при назначении цены на дополнительные услуги.

С учетом изложенных обстоятельств, возникает задача исследования благосостояния всех регионов. Ввиду большого количества субъектов в Российской Федерации задача исследования благосостояния каждого региона по отдельности становится трудоемкой. Для оптимизации данного процесса можно использовать методы кластеризации данных [1], чтобы получить группы, которые будут содержать регионы с примерно одинаковым уровнем жизни. Стоит разделить все категории следующим образом: «Авто», «Недвижимость», «Работа», «Услуги» и остальные категорий вместе взятые. Данное разделение обосно-

вано зависимостью цен на товары от уровня жизни в конкретном субъекте. Задача состоит в проведении нескольких процессов кластеризации для каждой категории отдельно, учитывая специфику товара в каждой из них.

В качестве основных данных для всех категорий объявлений будем использовать среднюю заработную плату [2] и численность населения [3] в соответствующем регионе.

Рассмотрим категорию «Авто». По причине того, что стоимость автомобиля не зависит от его местонахождения, стоит учитывать их количество в каждой области. Чем больше автомобилей, тем больше конкуренция. Соответственно возрастает и цена на услуги. Таким образом, в качестве дополнительного параметра будем использовать количество автомобилей, зарегистрированных в конкретной области [4].

В категории «Недвижимость» местонахождение объекта играет главную роль. Следовательно, чем дороже стоимость, например, квартиры, тем дороже услуги для объявления о ее продаже. Дополнительный параметр для данной категории – средняя стоимость квадратного метра жилья по области [5].

Цена на услуги для объявлений категории «Работа» должна зависеть от количества работодателей в конкретной области. В качестве дополнительного параметра используем количество

организаций, зарегистрированных в регионе [6].

В категория «Услуги» в роли дополнительного параметра будет использоваться объем платных услуг, оказанных населению области за год [7]. Подробнее с используемыми данными можно ознакомиться в Приложении 1.

2.2. Методы кластеризации

Все алгоритмы кластеризации делятся на иерархические и неиерархические [8]. Отличаются они по типу выходных данных. При использовании иерархического алгоритма мы получим результат в виде иерархии кластеров. В случае использовании алгоритма другого типа никакой иерархии у выходных данных не будет. Иерархические алгоритмы включают в себя дивизимные и агломеративные. Принцип работы дивизимного алгоритма основан на объединении всех объектов в общий кластер и последующем разделении этого кластера. Агломеративный алгоритм работает наоборот. Все объекты помещаются в собственные кластеры и только потом объединяются.

К неиерархическим алгоритмам относятся плотностные, модельные и итерационные. Принцип работы плотностного алгоритма заключается в представлении кластера как группы объектов, расположенных близко друг к другу. Модельный алго-

ритм выделяет в данных модели, которые и будут являться кластерами. Итеративные алгоритмы основаны на перераспределении объектов между кластерами. На них и остановимся подробнее. Одними из представителей итеративных алгоритмов являются k -means и c -means. Основное действие этих алгоритмов заключается в минимизации расстояния между объектами в кластерах.

Для начала выполнения алгоритма необходимо знать нужное количество кластеров. Выбор этого числа может зависеть от результатов прошлых исследований, теории или интуиции.

Описание алгоритма k -means

Минимизируемой функцией для k -means кластеризации является:

$$J = \sum_{k=1}^M \sum_{i=1}^N d^2(x_i, c_k),$$

$x_i \in X$ – объект (точка), $c_k \in C$ – центроид (центр кластера),
 $|X| = N$, $|C| = M$.

- Выбираются точки C . Сначала они будут являться центрами кластеров. Выбор центроидов может происходить случайно или вручную.
- Объекты перераспределяются по кластерам с помощью минимизации расстояния от объекта кластера до центра.

- На момент, когда объекты находятся в кластерах, происходит пересчет центров:

$$c_i = \frac{\sum_{i=1}^L x_i}{L},$$

где $x_i \in C_j$, а $|C_j| = L$.

- Центроиды считаются стабилизированными, если $c_i = c_i - 1$.

Выбор количества кластеров – важная часть процесса кластеризации. Если конкретного количества нет, то нужно поступать методом индукции. Провести несколько процессов и сравнить результаты. После получения результатов кластерного анализа нужно оценить, насколько кластеры отличаются друг от друга. Для такой оценки сравниваются средние значения для каждого из кластеров. При правильной кластеризации они должны сильно отличаться.

Достоинства k -means:

- простой
- быстрый
- понятный

Недостатки k -means:

- медленная работа с большими объемами данных

Описание алгоритма *c*-means

Этот алгоритм, в отличие от *k*-means, является нечетким методом кластеризации. Он основан на предположении, что определенный элемент принадлежит определенному кластеру с определенной долей вероятности. Минимизируемая функция:

$$J = \sum_{i=1}^N \sum_{k=1}^M u_{ij}^m d(x_i, c_k),$$

где также $x_i \in X$ – объект, $c_k \in C$ – центроид, $|X| = N$, $|C| = M$.

- Выбираем число классов M и меру m : $1 < m < \infty$, функцию расстояния $d(x, c)$ и критерий остановки поиска $0 < \varepsilon < 1$. Задаем U^0 – матрицу весов принадлежности определенной точки к определенному кластеру.

- Вычисляем центры кластеров:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$$

- Пересчитываем веса:

$$u_{ij} = \frac{1}{\sum_{k=1}^M \left(\frac{d(c_j, x_i)}{d(c_k, x_i)} \right)^{\frac{2}{m-1}}}$$

- Проверка условия $\|U^k - U^{k-1}\| < \varepsilon$. Если выполняется, то конец алгоритма.

Из описанных методов к нашей задаче наиболее подходит *k*-means. Его основные черты, которые послужат помощью в решении нашей задачи: четкость, простота и быстрая работа. Минусы данного метода не отразятся на результате ввиду того, что объем входных данных является небольшим.

2.3. Эмпирическая оценка стоимости платных услуг

Для решения основной задачи потребуется провести эмпирическую оценку функции спроса [9]. Для начала проводится опрос продавцов, которые зарегистрированы продолжительное время и имеют опыт продаж на сайте. Данную процедуру можно реализовать в личном кабинете пользователя. Опрос должен содержать следующие пункты: место жительства продавца, категория (подкатегория) товаров (должны быть доступны только те, в которых пользователь уже совершал сделки) и сумма, которую продавец был бы готов заплатить за каждую из предлагаемых сайтом услуг. Логично предположить, что не все пользователи согласятся пройти данный опрос и некоторые категории (подкатегории) определенных регионов не будут иметь статистических данных. Эта проблема решается наличием кластеров. Все данные будут собираться не для конкретного региона, а для определенного кластера. При этом, следует разделять

областные центры, рядовые города, поселки городского типа и села.

Получив результаты опроса, построим оценку функции спроса на платные услуги сайта. Далее определяем оптимальную цену на конкретную услугу с учетом доступности пользователю и максимальной прибыли компании.

2.4. Основные результаты и обсуждение

Написана программа в среде MATLAB, считывающая базу с данными для кластеризации и автоматически разделяющая столбцы для каждой категории объявлений. Кластеризация проводилась с помощью метода k -means, который реализован в пакете Statistics Toolbox для MATLAB [10]. Количество кластеров определялось эмпирическим путем, и в данном случае количество кластеров для всех категорий равно пяти. Код программы представлен в Приложении 2.

Результаты кластеризации для категории «Авто»

- 1-й кластер: 9 регионов
- 2-й кластер: 47 регионов
- 3-й кластер 24 региона

- 4-й кластер: 3 региона
- 5-й кластер: 1 регион

Результаты кластеризации для категории «Недвижимость»

- 1-й кластер: 10 регионов
- 2-й кластер: 50 регионов
- 3-й кластер 21 региона
- 4-й кластер: 2 региона
- 5-й кластер: 1 регион

Результаты кластеризации для категории «Работа»

- 1-й кластер: 57 регионов
- 2-й кластер: 15 регионов
- 3-й кластер 9 региона
- 4-й кластер: 2 региона
- 5-й кластер: 1 регион

Результаты кластеризации для категории «Услуги»

- 1-й кластер: 55 регионов

- 2-й кластер: 16 регионов
- 3-й кластер 9 региона
- 4-й кластер: 3 региона
- 5-й кластер: 1 регион

Результаты кластеризации для остальных категорий

- 1-й кластер: 9 регионов
- 2-й кластер: 49 регионов
- 3-й кластер 22 региона
- 4-й кластер: 3 региона
- 5-й кластер: 1 регион

Подробные результаты представлены в Приложении 3.

Оценка функции спроса на услугу «Премиум-размещение»

Для примера возьмем г. Санкт-Петербург (четвертый кластер категории «Авто»). Пусть, для определенности, выборка из результатов опроса для услуги «Премиум-размещение» в подкатегории «Автомобили» выглядит следующим образом:

1100,	1300,	1000,	1300,	900,	1000,	900,	900,	1000,	1000,
1500,	800,	1000,	1200,	900,	900,	1000,	1100,	1000,	1100,
700,	1400,	1600,	1000,	800,	1300,	800,	1100,	1400,	900,
1000,	700,	700,	1200,	900,	1300,	900,	700,	1100,	1000,
1000,	1200,	1200,	800,	1100,	1100,	1000,	1200,	1300,	1200.

Следующим шагом оценим функцию спроса. Для начала упорядочиваем максимально допустимые для опрашиваемых цены на услугу. Результаты представлены во втором столбце. В третьем столбцы указано количество повторений той или иной цены. Далее строим функцию спроса, которая располагается в четвертом столбце. Из результатов опроса замечаем, что никто из опрошенных данную услугу не приобретет по цене, превышающей 1600. При цене, равной 1600, появится только один покупатель. При снижении цены до 1500, услугу смогут позволить себе приобрести уже двое пользователей. Аналогичным образом заполняется весь столбец. В пятом столбце отражена прибыль от продаж услуги по определенной цене.

№, i	Цена, p_i	Повторы, N_i	Спрос, $D(p_i)$	Прибыль, $p_i D(p_i)$
1	700	4	50	35000
2	800	4	46	36800
3	900	8	42	37800
4	1000	12	34	34000
5	1100	7	22	24200
6	1200	6	15	18000
7	1300	5	9	11700
8	1400	2	4	5600
9	1500	1	2	3000
10	1600	1	1	1600

Ввиду отсутствия издержек, оптимальную цену услуги определить достаточно просто. В нашем случае это та цена, при которой мы получим максимальную прибыль, т. е. 900.

Аналогичным образом рассчитаем оптимальную цену на остальные услуги для категории «Авто» подкатегории «Автомобили».

Оценка функции спроса на услугу «VIP-объявление»

№, i	Цена, p_i	Повторы, N_i	Спрос, $D(p_i)$	Прибыль, $p_i D(p_i)$
1	100	3	50	5000
2	200	3	47	9400
3	300	8	44	13200
4	400	9	36	14400
5	500	10	27	13500
6	600	11	17	10200
7	700	3	6	4200
8	800	2	3	2400
9	900	1	1	900

По аналогии с предыдущими расчетами замечаем, что максимальная прибыль будет при цене 400. Это и есть оптимальная цена на данную услугу.

Оценка функции спроса на услугу «Поднятие объявление в поиске»

№, i	Цена, p_i	Повторы, N_i	Спрос, $D(p_i)$	Прибыль, $p_i D(p_i)$
1	0	2	50	0
2	100	10	48	4800
3	200	20	38	7600
4	300	13	18	5400
5	400	5	5	2000

Данная услуга интересна обеим сторонам по цене 200.

Оценка функции спроса на услугу «Выделение объявления»

№, i	Цена, p_i	Повторы, N_i	Спрос, $D(p_i)$	Прибыль, $p_i D(p_i)$
1	0	3	50	0
2	100	11	47	4700
3	200	24	36	7200
4	300	9	12	3600
5	400	3	3	1200

Оптимальная цена этой услуги совпадает с ценой на услугу «Поднятие объявления в поиске», т. е. 200.

Глава 3. Математическая модель репутации пользователей

На данный момент сайт Avito не участвует в процессе сделки между пользователями и не является гарантом. Данный факт значительно отличает сайт от таких популярных торговых площадок, как Ebay, Amazon, Taobao. В 2015 году компания приобрела сервис доставки Checkout, который является агрегатором доставок для интернет-магазинов, позволяющий выбирать из нескольких способов доставки, таких как Почта России, SPSR Express, DPD, PickPoint и автоматически рассчитывает стоимость доставки до клиента.

3.1. Постановка задачи

Учитывая наличие сервиса доставки и возможность с помощью него контролировать процесс сделки, появится необходимость во введении шкалы репутации продавцов. Следовательно, каждый продавец будет стараться заработать показатель репутации, стремящийся к 100%, т. к. рейтинг показывает надежность [11] и является одним из главных критериев выбора покупателем того или иного продавца [12].

3.2. Метод оценки репутации

Оставить отзыв покупатель сможет после совершения сделки в собственном личном кабинете на сайте. Должны быть доступны два варианта отзыва: «Положительный» и «Отрицательный». При расчете они должны иметь значения «1» и «-1» соответственно. Последний подразумевает поле для подробного описания жалобы. Данные отзывы должны обрабатываться соответствующим отделом сотрудников.

К сожалению, не исключены случаи искусственного повышения или понижения рейтинга с целью введения в заблуждение потенциального покупателя и проведения в отношении него каких-либо мошеннических действий или устранения конкурентов. Чтобы снизить количество искусственных оценок репутации, нужно ввести сбор, который уплачивается новым пользователем при регистрации на сайте. Данный взнос должен возвращаться на кошелек пользователя спустя некоторое время или после его первой продажи на сайте [13]. Кроме того, оставить отрицательный отзыв о конкретном продавце каждый пользователь может только один раз. Это объясняется тем, что покупатель не будет совершать сделку с человеком, с которым уже был неприятный опыт сотрудничества. Также стоит не принимать первую жалобу на продавца и отклонить ее

до появления аналогичной уникальной жалобы. Данные условия помогут минимизировать процент искусственных отзывов и предоставят потенциальным покупателям реальный рейтинг продавца.

Для расчета репутации нужно учитывать стоимость товара. Таким образом, вес отрицательного отзыва для более дорогого лота, является большим, относительно отрицательного отзыва для дешевого товара. Это позволит избежать случаев, когда продавец искусственно увеличивает свою репутацию посредством продажи дешевых товаров и совершать в дальнейшем мошеннические действия, предлагая дорогие товары. Следует установить границы ценовых диапазонов и производить расчет репутации отдельно для каждой из них.

3.3. Основные результаты и обсуждение

При совершении сделки покупатель оставляет отзыв о продавце. После этого рассчитывается общий рейтинг для всех ценовых диапазонов за текущий месяц. В данном моменте должны учитываться цены на проданные товары. Если рейтинг в i -ом диапазоне < 0.9 , то вес отзыва по данному диапазону уменьшается в i раз. Это объясняется тем, что десятая часть отрицательных отзывов может быть допустимой на случаи недо-

понимания между продавцом и покупателем. Если репутация в i -ом диапазоне < 0 , то данное значение увеличивается в i раз.

Формула для расчета выглядит следующим образом:

$$M_i = \sum_{j=1}^3 \frac{1}{3n} \left(\frac{c_{ij}}{k_{ij}} \right),$$

c_{ij} – сумма оценок продавца в j -ом диапазоне цен, за i -ый месяц, k_{ij} – количество оценок в j -ом диапазоне цен, за i -ый месяц, а значение n зависит от N_j :

$$n = \begin{cases} 1, & N_j \in [0.9, 1] \\ j, & N_j \in [0, 0.9) \\ \frac{1}{j}, & N_j \in (-\infty, 0) \end{cases}$$

Заключительным этапом расчета рейтинга является учет показателей за предыдущие пять месяцев. Кроме того, нужно учитывать то, что новые показатели репутации важнее, чем прошлые. Поэтому нужно увеличить вес отрицательных отзывов, в зависимости от месяца. Таким образом, значение репутации продавца, которая будет видна зарегистрированному пользователю, рассчитывается за последние полгода и имеет вид:

$$F = \sum_{i=1}^6 \frac{1}{6m} M_i,$$

где значение m зависит от M_i :

$$n = \begin{cases} 1, & i \in [1, 2] \\ 2, & i \in [3, 4]; M_i \in [0, 0.9) \\ 3, & i \in [5, 6]; M_i \in [0, 0.9) \\ \frac{1}{2}, & i \in [3, 4]; M_i \in (-\infty, 0) \\ \frac{1}{3}, & i \in [5, 6]; M_i \in (-\infty, 0) \end{cases} ,$$

а M_i находится уже по известной формуле:

$$M_i = \sum_{j=1}^3 \frac{1}{3n} \left(\frac{c_{ij}}{k_{ij}} \right) .$$

Примеры работы модели

Рассмотрим различные сценарии сделок на сайте и изменение репутации продавца.

Допустим, что все оценки продавца в текущем месяце были положительными для 1-го и 2-го диапазона и, соответственно, его рейтинг был максимальным. Следующей его оценкой стала «-1» для 3-го диапазона.

$i \backslash k$	1	2	3	4
1	1	1	1	1
2	1	1	1	1
3	-1			

Репутация за предыдущие пять месяцев выглядит следующим образом:

j	1	2	3	4	5	6
M_j	0.9	0.9	0.9	0.9	0.9	

В результате получим, что за текущий месяц репутация пользователя изменилась с максимальной до -0.33 , а общая репутация снизилась с 0.92 до 0.59 .

Пусть, продавец совершил еще одну сделку с ценой, входящей в 3-ий диапазон.

$i \backslash k$	1	2	3	4
1	1	1	1	1
2	1	1	1	1
3	-1	1		

Тогда рейтинг текущего месяца возрастет до 0.66 , а общий станет -0.79 .

Пусть продавец имел положительные оценки в 1-ом диапазоне и, соответственно, максимальный рейтинг.

$i \backslash k$	1	2	3	4
1	1	1	1	1
3	-1			

j	1	2	3	4	5	6
M_j	0.9	0.9	0.9	0.9	0.9	

При получении отрицательной оценки в 3-ем диапазоне, его общая репутация снизится с 0.92 до 0.5.

Заключение

В работе предложен способ установления стоимости платных услуг на сайте Avito с помощью метода кластеризации данных и эмпирической оценки функции спроса. В среде MATLAB реализована программа, самостоятельно проводящая соответствующие процессы. Учтено благосостояние субъектов Российской Федерации, дополнительные характеристики и специфика категорий объявлений, в которых доступно размещение. Как и следовало ожидать, в самом богатом кластере оказался г. Москва. В четвертом кластере, практически по всех категориях оказались г. Санкт-Петербург, Московская область и Краснодарский край. Построена эмпирическая оценка функции спроса на платные услуги сайта Avito и определена оптимальная цена на них.

Кроме того, построена математическая модель репутации пользователей, которая должным образом реагирует на некорректное ведение сделок продавцом, на искусственное завышение или занижение репутации другими пользователями, с целью совершения мошеннических действий или устранения конкурентов. Приведены примеры работы данной модели, которые показывают ее поведение в различных ситуациях.

Список литературы

1. Буре В. М., Екимов А. В., Свиркин М. В. Имитационная модель формирования профиля мнений внутри коллектива // Вестник Санкт-Петербургского университета. Сер. 10: Прикладная математика. Информатика. Процессы управления. 2014. № 3. С. 93–98.
2. Среднемесячная номинальная начисленная заработная плата на одного работника по полному кругу организаций // ЕМИСС [Электронный ресурс]: URL:<https://fedstat.ru/indicator/33433> (дата обращения: 02.02.16).
3. Численность населения по итогам Всероссийской переписи населения // ЕМИСС [Электронный ресурс]: URL:<https://fedstat.ru/indicator/37427> (дата обращения: 02.02.16).
4. Обеспеченность россиян автомобилями за год увеличилась на 3,2% // Страховая компания – Группа «АльфаСтрахование»: страховые продукты для частных лиц и корпоративных клиентов [Электронный ресурс]: URL:http://www.alfastrah.ru/news/index.php?ELEMENT_ID=721580&sphrase_id=427814 (дата обращения: 02.02.16).

5. Средняя цена одного кв.м общей площади квартир на рынке жилья // ЕМИСС [Электронный ресурс]: URL:<https://fedstat.ru/indicator/31452> (дата обращения: 02.02.16).
6. Количество организаций по данным государственной регистрации // ЕМИСС [Электронный ресурс]: URL:<https://fedstat.ru/indicator/42930> (дата обращения: 02.02.16).
7. Объем платных услуг населению // ЕМИСС [Электронный ресурс]: URL:<https://fedstat.ru/indicator/31280> (дата обращения: 02.02.16)
8. Charu C. Aggrwal, Chandan K. Reddy Data Clustering: Algorithms and Applications, Chapman & Hall/CRC, 2013. 652 p.
9. Орлов А. И. Эконометрика. Учебник. М.: Экзамен, 2002. 576 с.
10. MathWorks – Makers of MATLAB and Simulink // k-means clustering [Электронный ресурс]: URL:<https://www.mathworks.com/help/stats/kmeans.html> (дата обращения: 10.02.16).
11. Wu F., Li H.-H., Kuo Y.-H. Reputation evaluation for choosing a trustworthy counterparty in C2C e-commerce //

Electronic Commerce Research and Applications, 2011. Vol. 18, № 3, P. 428-436.

12. Chang J.- S., Wong H.- J. Selecting appropriate sellers in online auctions through a multi-attribute reputation calculation method // Electronic Commerce Research and Applications, 2011. Vol. 10, № 2, P. 144-154.
13. Panayotis F. A novel reputation-based model for e-commerce // Operational Research, 2013. Vol. 13, № 1, P. 113-138.

Приложение

Приложение 1. Данные для кластеризации

Subject	Salary	Auto	Realty	Job	Services	Popular
Алтайский к.	19456	583161	43817	55771	72008070	2419755
Амурская о.	32397	186773	56138	16162	38510697	830103
Архангельская о.	33125	267622	62801	24515	57112136	1227626
Астраханская о.	24576	254538	44484	18737	34874271	1010073
Белгородская о.	23895	419912	50973	35745	70453405	1532526
Брянская о.	20911	240305	32458	22548	42682044	1278217
Владимирская о.	22581	319056	44412	35441	64011050	1443693
Волгоградская о.	22828	605557	40460	55485	120534931	2610161
Вологодская о.	26749	316243	41234	43107	50738910	1202444
Воронежская о.	24001	595522	48315	57481	99563753	2335380
г. Москва	61208	3416540	192292	1138560	1324155941	11503501
г. СПб	40697	1439472	83263	357242	386443870	4879566
Еврейская АО	29439	40961	43099	3362	7364349	176558
Забайкальский к.	29319	259063	49866	16662	38839490	1107107
Ивановская о.	20592	186851	45575	34120	36921299	1061651
Иркутская о.	31408	650905	43619	66204	84474425	2428750
Калининградская о.	26639	299516	50280	53061	41851717	941873
Калужская о.	28248	324509	55878	27990	41290776	1010930
Камчатский к.	53167	145580	61532	11519	27104609	322079
Кемеровская о.	26809	577495	40377	51068	91909532	2763135
Кировская о.	20978	288382	45375	38841	48021746	1341312
Костромская о.	20867	154874	42222	17778	22608490	667562
Краснодарский к.	25777	1426875	57058	139913	360831938	5226647
Красноярский к.	34178	743813	46741	78090	130323566	2828187
Курганская о.	21172	234988	39778	17445	27001010	910807
Курская о.	23099	281770	42101	24230	44706939	1127081
Ленинградская о.	31851	465271	61457	38185	57996751	1716868
Липецкая о.	23133	305113	47769	21692	50959322	1173513
Магаданская о.	62152	42232	51810	5296	13137037	156996

Московская о.	38598	2291724	76684	248852	427558186	7095120
Мурманская о.	43378	224305	50123	21552	54787533	795409
Ненецкий АО	65816	8208	82717	1114	2512852	42090
Нижегородская о.	25497	741574	66698	97764	149205526	3310597
Новгородская о.	25225	171210	46645	16384	31720530	634111
Новосибирская о.	27214	711798	55130	142813	129952090	2665911
Омская о.	26205	526059	44873	49088	77460182	1977665
Оренбургская о.	23469	585525	43968	40936	80014487	2033072
Орловская о.	20885	212472	35902	16609	28261033	786935
Пензенская о.	22392	310506	41633	27590	43357959	1386186
Пермский к.	27102	534961	40965	77199	130983732	2635276
Приморский к.	32431	1070204	85047	67834	138156738	1956497
Псковская о.	21004	210781	38332	15992	21039190	673423
Р. Адыгея	20945	110879	38733	7500	11243300	439996
Р. Алтай	22598	41234	42921	6484	3934828	206168
Р. Башкортостан	24988	965133	54922	88888	218847512	4072292
Р. Бурятия	27739	179824	43962	20136	38748437	972021
Р. Дагестан	18194	436537	45641	33869	96395555	2910249
Р. Ингушетия	20993	47441	35438	5347	9582868	412529
Р. К.-Балкария	20323	166828	39648	12646	24030151	859939
Р. Калмыкия	19341	60212	40564	5547	4608485	289481
Р. К.-Черкессия	19746	69767	44006	6964	11757972	477859
Р. Карелия	29371	200143	49087	23520	27318383	643548
Р. Коми	40222	220791	58158	21103	45734516	901189
Р. Крым	17912	649319	52820	22000	8624208	1908322
Р. Марий Эл	20473	113523	41097	15788	19613429	696459
Р. Мордовия	20342	171960	41674	16505	22187369	834755
Р. Саха	51111	165825	82462	26879	67744867	958528
Р. С. Осетия	20311	151152	40030	11862	23268023	712980
Р. Татарстан	28294	988273	58935	118995	229759711	3786488
Р. Тыва	27507	50193	50342	3815	5274981	307930
Р. Удмуртия	23421	345362	45743	38585	52494730	1521420
Р. Хакасия	29085	148008	46198	11727	15786830	532403

Р. Чечня	21452	143396	42161	9690	37771099	1268989
Р. Чувашия	20854	217782	42915	25372	39945179	1251619
Ростовская о.	23818	1112274	53667	89297	171447339	4277976
Рязанская о.	24280	337001	46524	34068	37908510	1154114
Самарская о.	25884	868194	54103	110587	144819480	3215532
Саратовская о.	22012	645604	40999	50292	78937770	2521892
Сахалинская о.	54896	140428	79194	17485	44552489	497973
Свердловская о.	29492	1044353	60364	168241	281070867	4297747
Смоленская о.	22279	267081	36361	26946	30038173	985537
Ставропольский к.	22597	660349	34212	56821	130278576	2786281
Тамбовская о.	20757	291562	41476	17813	42424637	1091994
Тверская о.	23866	369476	50692	36057	42022753	1353392
Томская о.	32042	264991	49448	36226	42019997	1047394
Тульская о.	25873	414898	50236	35356	59850535	1553925
Тюменская о.	34125	1015331	59973	46425	66577249	3395755
Ульяновская о.	21272	294758	43875	29023	44894484	1292799
Хабаровский к.	36781	255335	61771	44561	118595549	1343869
Х.-Мансийский АО	57976	458141	58937	41838	92177389	1532243
Челябинская о.	27683	879483	40926	108092	134025764	3476217
Чукотский АО	76285	3385	32924	1208	4228570	50526
Я.-Ненецкий АО	74489	131249	60465	11527	37218937	522904
Ярославская о.	25434	258311	48730	45347	46011792	1272468

Приложение 2. Реализация процессов кластеризации в MATLAB

```
clear, clc;
Universal = xlsread('Data for clustering.xls', 'Лист2')
Universal(:,1) = [ ]
Universal(:,1) = [ ]

n = size(Universal, 1)
m = size(Universal, 2)

Clusters = cell(84,1)

Auto = zeros(n,1)
Realty = zeros(n,1)
Job = zeros(n,1)
Services = zeros(n,1)
Others = zeros(n,1)

for i = 1:m
    if Universal(85,i) == 0
        Auto = [Auto, Universal(:,i)]
        Realty = [Realty, Universal(:,i)]
        Job = [Job, Universal(:,i)]
        Services = [Services, Universal(:,i)]
        Others = [Others, Universal(:,i)]
    end

    if Universal(85,i) == 1
        Auto = [Auto, Universal(:,i)]
    end

    if Universal(85,i) == 2
        Realty = [Realty, Universal(:,i)]
    end
end
```

```

    if Universal(85,i) == 3
        Job = [Job, Universal(:,i)]
    end

    if Universal(85,i) == 4
        Services = [Services, Universal(:,i)]
    end

end

Auto(:,1) = [ ]
Realty(:,1) = [ ]
Job(:,1) = [ ]
Services(:,1) = [ ]
Others(:,1) = [ ]

Auto(n,:) = [ ]
Realty(n,:) = [ ]
Job(n,:) = [ ]
Services(n,:) = [ ]
Others(n,:) = [ ]

opts = statset('Display','final');

k = 5;
startPositions1 = [0.0 0.0 0.0; 0.2 0.2 0.2;
    0.4 0.4 0.4; 0.6 0.6 0.6; 1.0 1.0 1.0];
startPositions2 = [0.0 0.0 0.0; 0.2 0.2 0.2;
    0.4 0.4 0.4; 0.6 0.6 0.6; 1.0 1.0 1.0];
startPositions3 = [0.1 0.1 0.1; 0.2 0.2 0.2;
    0.3 0.3 0.3; 0.4 0.4 0.4; 1.0 1.0 1.0];
startPositions4 = [0.1 0.1 0.1; 0.2 0.2 0.2;
    0.3 0.3 0.3; 0.4 0.4 0.4; 1.0 1.0 1.0];
startPositions5 = [0.0 0.0; 0.2 0.2; 0.4 0.4;
    0.6 0.6; 1.0 1.0];

```

```

Clusters1 = kmeans(Auto,k,...
                    'Distance','city',...
                    'Options',opts,...
                    'Start',startPositions1 ...
                );
Clusters2 = kmeans(Realty,k,...
                    'Distance','city',...
                    'Options',opts,...
                    'Start',startPositions2 ...
                );
Clusters3 = kmeans(Job,k,...
                    'Distance','city',...
                    'Options',opts,...
                    'Start',startPositions3 ...
                );
Clusters4 = kmeans(Services,k,...
                    'Distance','city',...
                    'Options',opts,...
                    'Start',startPositions4 ...
                );
Clusters5 = kmeans(Others,k,...
                    'Distance','city',...
                    'Options',opts,...
                    'Start',startPositions5 ...
                );
Clusters = [Clusters1, Clusters2, Clusters3, Clusters4, Clusters5]
xlswrite('Clusters.xls', Clusters, 'C2:G85')

```

Приложение 3. Результаты кластеризации

Subject	Auto	Realty	Job	Services	Others
Алтайский край	3	2	2	2	3
Амурская область	2	2	1	1	2
Архангельская область	2	2	1	1	2
Астраханская область	2	2	1	1	2
Белгородская область	2	2	1	1	2
Брянская область	2	2	1	1	2
Владимирская область	2	2	1	1	2
Волгоградская область	3	3	2	2	3
Вологодская область	2	2	1	1	2
Воронежская область	3	3	2	2	3
г. Москва	5	5	5	5	5
г. Санкт-Петербург	4	4	4	4	4
Еврейская АО	2	2	1	1	2
Забайкальский край	2	2	1	1	2
Ивановская область	2	2	1	1	2
Иркутская область	3	3	2	2	3
Калининградская область	2	2	1	1	2
Калужская область	2	2	1	1	2
Камчатский край	1	1	1	1	1
Кемеровская область	3	3	2	2	3
Кировская область	2	2	1	1	2
Костромская область	2	2	1	1	2
Краснодарский край	4	3	3	4	4
Красноярский край	3	3	2	3	3
Курганская область	2	2	1	1	2
Курская область	2	2	1	1	2
Ленинградская область	3	3	1	2	3
Липецкая область	2	2	1	1	2
Магаданская область	1	1	1	1	1
Московская область	4	4	4	4	4
Мурманская область	1	1	1	1	1

Ненецкий АО	1	1	1	1	1
Нижегородская область	3	3	3	3	3
Новгородская область	2	2	1	1	2
Новосибирская область	3	3	2	2	3
Омская область	3	2	2	2	3
Оренбургская область	3	2	2	2	2
Орловская область	2	2	1	1	2
Пензенская область	2	2	1	1	2
Пермский край	3	3	2	2	3
Приморский край	3	3	2	2	3
Псковская область	2	2	1	1	2
Республика Адыгея	2	2	1	1	2
Республика Алтай	2	2	1	1	2
Республика Башкортостан	3	3	3	3	3
Республика Бурятия	2	2	1	1	2
Республика Дагестан	3	3	2	2	3
Республика Ингушетия	2	2	1	1	2
Республика Кабардино-Балкария	2	2	1	1	2
Республика Калмыкия	2	2	1	1	2
Республика Карачаево-Черкессия	2	2	1	1	2
Республика Карелия	2	2	1	1	2
Республика Коми	2	1	1	1	2
Республика Крым	3	2	2	1	2
Республика Марий Эл	2	2	1	1	2
Республика Мордовия	2	2	1	1	2
Республика Саха (Якутия)	1	1	1	1	1
Республика Северная Осетия	2	2	1	1	2
Республика Татарстан	3	3	3	3	3
Республика Тыва	2	2	1	1	2
Республика Удмуртия	2	2	1	1	2
Республика Хакасия	2	2	1	1	2
Республика Чечня	2	2	1	1	2
Республика Чувашия	2	2	1	1	2

Ростовская область	3	3	3	3	3
Рязанская область	2	2	1	1	2
Самарская область	3	3	3	3	3
Саратовская область	3	2	2	2	3
Сахалинская область	1	1	1	1	1
Свердловская область	3	3	3	3	3
Смоленская область	2	2	1	1	2
Ставропольский край	3	3	2	2	3
Тамбовская область	2	2	1	1	2
Тверская область	2	2	1	1	2
Томская область	2	2	1	1	2
Тульская область	2	2	1	1	2
Тюменская область	3	3	3	3	3
Ульяновская область	2	2	1	1	2
Хабаровский край	2	3	1	2	2
Ханты-Мансийский АО	1	1	1	2	1
Челябинская область	3	3	3	3	3
Чукотский АО	1	1	1	1	1
Ямало-Ненецкий АО	1	1	1	1	1
Ярославская область	2	2	1	1	2